New York City Airbnb: Price Prediction

Dataset Background & Structure

The New York City Airbnb is a public dataset and a part of Airbnb. This dataset displays the listing activities of New York City for 2019. the The dataset was downloaded as a CSV and it contains 16 columns and 48,895 rows for which all appear to be unique. It contains neighbourhood groups, location coordinates, room types, price, minimum nights, review related data, and availability.

Quality Issues

Some of the quality issues that we can see on a quick analysis are:
- A number of missing data in review-related columns
- Features with different value range

Exploratory Data Analysis Goals

The dataset provides various information about the large number of existing listings. We will explore data in order to mainly discover relationships between provided features and price and make prediction about price.

First, before analysis, we will clean data; we will remove columns with many missing values, and make value range consistent in order to ensure each feature with numeric value have the same impact.

Then, we will perform EDA(exploratory Data Analysis) to understand the main characteristics of data, uncover the relationships between listing features, and extract the important features affecting price.

Once the analysis process is complete, the possible use for the dataset would be to provide price reference or insight to New-York-based hosts new to Airbnb or any type of accommodation search sites and applications.